

# Studying co-movements in large multivariate models without multivariate modelling

Citation for published version (APA):

Cubadda, G., Hecq, A. W., & Palm, F. C. (2007). *Studying co-movements in large multivariate models without multivariate modelling*. METEOR, Maastricht University School of Business and Economics. METEOR Research Memorandum No. 032 <https://doi.org/10.26481/umamet.2007032>

## Document status and date:

Published: 01/01/2007

## DOI:

[10.26481/umamet.2007032](https://doi.org/10.26481/umamet.2007032)

## Document Version:

Publisher's PDF, also known as Version of record

## Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

## General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

## Take down policy

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Gianluca Cubadda, Alain Hecq, Franz.C. Palm

Studying Co-movements in Large Multivariate  
Models Without Multivariate Modelling

RM/07/032

JEL code: C32



Maastricht research school of **E**conomics  
of **T**Echnology and **O**Rganizations

Universiteit Maastricht  
Faculty of Economics and Business Administration  
P.O. Box 616  
NL - 6200 MD Maastricht

phone : ++31 43 388 3830  
fax : ++31 43 388 4873

# STUDYING CO-MOVEMENTS IN LARGE MULTIVARIATE MODELS WITHOUT MULTIVARIATE MODELLING

Gianluca Cubadda                      Alain Hecq                      Franz C. Palm\*  
Università di Roma "Tor Vergata"      Maastricht University      Maastricht University

August 22, 2007

## Abstract

We propose an approach for checking the data admissibility of non-stationary multivariate time series models (VAR or VARMA) through that of their implied individual ARIMA specifications. In particular we show that the presence of different kinds of common cyclical features restrictions, leading to reduced rank in the short-run dynamics, explains why we identify parsimonious univariate ARIMA models in applied research, a paradox that the profession had difficulties to explain. We develop a new strategy for studying interactions between variables prior to possibly modelling them in a multivariate setting. Indeed, we provide tools to study features of individual time series with the aim to infer features of the complete system, as individual series keep a print of the system as a whole. The similarity of the autoregressive roots will be informative about co-movements existing in a vector autoregressive model as well as convergence between series for different economies. It will allow us to forecast series, to build business cycle indices, to unravel trends from cycles in a way that is consistent with the full multivariate system. Our results justify both the use of an homogeneous panel with heterogeneous cross-correlated vector moving average (VMA) errors and a factor structure, and the cross-sectional aggregation of ARIMA series. The advantages of our approach are many: 1) determining co-movements also when we cannot work with a complete system, 2) enhancing the accuracy of forecasts, 3) the ease of its implementation in complex situations, 4) the potential empirical studies in many fields.

JEL: C32

Keywords: Interactions, time series, co-movements, ARIMA, cointegration, common cycles, panel data

---

\*Corresponding author: Franz C. Palm, Maastricht University, Department of Quantitative Economics, P.O.Box 616, 6200 MD Maastricht, The Netherlands. Email: f.palm@ke.unimaas.nl.

# 1 Introduction

The analysis of interactions between individual phenomena is a key objective of the scientific research. A huge collection of data (national accounts, unemployment for hundreds of categories, prices for thousands of goods, environmental data on CO2 or on climate, health care, education...) for virtually all countries over the world is almost instantaneously available at a relatively low cost. At the European level, starting for six countries, Eurostat now collects and analyzes hundreds of variables for 27 member states. The IMF or the Worldbank publish statistical information for 150 countries. Hence, the task of researchers is to provide methods to study interactions between large sets of data.

One strategy to capture interactions and to forecast a large set of time series is to develop models to study movements in many variables by imposing restrictions. An example is the dynamic factor model approach by e.g. Forni et al. (2000) and Stock and Watson (2002) but the recent literature on dynamic macro-panels falls also in this category. Alternatively partial systems are built for country specific analyses or for a specific variable such that the GDP for a number of countries. A well known framework to simultaneously analyze a limited number of time series is the vector autoregressive model (VAR hereafter). In VARs, allowing all variables to be endogenous is only implementable for small systems, that is to say, in general up to six series if one has at least quarterly observations and enough data points. Trying to cure the dimensionality problem in VARs, different approaches have been proposed such as *inter alia* Bayesian analyses, simulation based techniques, separability procedures, automatic selection by deleting non significant coefficients and reduced rank regressions.

In this paper we take another route. Instead of designing a system which captures interactions we show how to analyze individual series to extract information on a possibly large system because individual series keep a print of the system as a whole. Consequently, we propose a new strategy that allows to study interactions between variables, but without modelling those directly in a multivariate framework. However, before implementing this strategy, we start to explain in details our view on the compatibility between univariate and multivariate models. More precisely, this paper first extends in Section 2 results of Zellner and Palm (1974). We show that using VARs with additional restrictions coming from short-run co-movements (see *inter alia* Engle and Kozicki (1993), Vahid and Engle (1993)), we give an alternative explanation to a well known paradox in time series: on the one hand, small multivariate systems theoretically imply non parsimonious individual ARMA processes while on the other hand estimated univariate models need only few parameters. This finding that the presence of co-movements implies that the univariate ARMA schemes are of low order will be used in a new strategy for the validation of multivariate models through the implied univariate models. This is illustrated with the relationship between quarterly growth rates of the industrial production indexes in the US and Canada. The tools we propose have further implications that we can exploit on moving average components. Section 3 proposes an estimation strategy we evaluate in a small Monte Carlo experiment. In particular we propose to work on cross sectional aggregates to compute the common autoregressive parameters. Indeed, an homogeneous panel framework with common autoregressive coefficients is the by-product of a large VAR model. Section 4 focuses on the link between the growth rates of the GDP among Latin American economies. A final section concludes and gives our plans and hints for

further research.

## 2 Multivariate and univariate time series

### 2.1 The paradox

The literature on the link between multivariate time series and the behavior of univariate variables stresses the fact that univariate ARIMA analyses could provide tools for the diagnostic of VAR (or VARMA) models. Important contributions often quoted in this area are the monograph by Quenouille (1957) and papers by Zellner and Palm (1974, 1975, 2004), Palm and Zellner (1980), Palm (1977) or Maravall and Mathis (1994). Some textbooks devote a few pages to that issue (e.g. Franses (1998, 198-199) or Lütkepohl (2005, 494-495)). More formally, let us first consider the stationary<sup>1</sup> multiple time series process for an  $n$  vector of time series  $z_t = (z_{1t}, \dots, z_{nt})'$  and without deterministic terms for simplicity:

$$\Phi(L)z_t = \Theta(L)\varepsilon_t, \quad t = 1, \dots, T, \quad (1)$$

where  $\Phi(L)$  and  $\Theta(L)$  are finite order polynomial coefficient matrices with the usual lag operator  $L$  such that  $Lz_t = z_{t-1}$ . The sequence of  $\varepsilon_t$  is a multivariate white noise process with each  $\varepsilon_t \sim N(0, \Sigma)$ . The VARMA in (1) encompasses several useful specifications. For example, and using additional normalization rules on contemporaneous matrices: if  $\Phi(L) = I$  we have the vector moving average representation or VMA; if  $\Theta(L) = I$  we have a VAR( $p$ ). Let us commence the analysis with the VAR of order  $p$ , denoted VAR( $p$ ), one of the cornerstone specification in empirical macroeconometrics:

$$\Phi(L)z_t = \varepsilon_t. \quad (2)$$

Following Zellner and Palm (1974), the univariate representation of elements of  $z_t$  can be obtained by premultiplying both sides of (2) by  $\Phi(L)^{adj}$ , the adjoint matrix (or the adjugate) associated with  $\Phi(L)$ , in order to obtain the "final equations" representation

$$\det[\Phi(L)]z_t = \Phi(L)^{adj}\varepsilon_t, \quad (3)$$

where the determinant  $\det[\Phi(L)]$  is a scalar finite order polynomial in  $L$ . This means that each series is a finite order ARMA( $p^*, q^*$ ), with the same lag structure and the same coefficients for the autoregressive part for every series, although the system was a finite order VAR( $p$ ). For instance for the  $i$ th element of  $z_t$  we have

$$\det[\Phi(L)]z_{it} = \Phi_{i.}(L)^{adj}\varepsilon_t = \theta_i(L)u_{it}, \quad (4)$$

---

<sup>1</sup> The generalisation to non-stationarity processes will be considered later in this section.

$\Phi_i(L)^{adj}$  denoting the  $i$ th row of the matrix  $\Phi(L)^{adj}$ ,  $\theta_i(L)$  is a scalar polynomial in  $L$  and  $u_{it}$  is a scalar white noise (innovation) with respect to the past of  $z_{it}$ . This recognition and the compatibility of these  $p^*$  and  $q^*$  with  $n$  and  $p$  of the  $\text{VAR}(p)$  is the first step of a methodology that has been developed in Zellner and Palm (1974, 1975) as a general modelling strategy called SEMTSA (*Structural Economic Modelling Time Series Analysis*). In the first stage of that framework, information from univariate schemes is used to restrict the dynamics of the structural model. However, given the conditions associated with that first diagnostic checking, one must often face a paradoxical feature: a small order VAR with few series already generates univariate ARMA with large  $p^*$  and  $q^*$ , a finding that is rejected when tested on economic data where one usually finds quite parsimonious models with low order autoregressive and moving average polynomials.

Indeed, an  $n$  dimensional  $\text{VAR}(p)$  would imply **at most** individual  $\text{ARMA}(np, (n-1)p)$  processes. This well known result is simply due to the fact that  $\det[\Phi(L)]$  contains by construction up to  $L^{np}$  terms and the adjoint matrix is a collection  $(n-1) \times (n-1)$  cofactor matrices, each of the matrix elements can contain the terms  $1, L, \dots, L^p$ . But whatever the simplicity of that result, this leads to implausible models for most observed economic time series.<sup>2</sup> This is illustrated in the following numerical example.

**Example 1** *Let us consider the bivariate  $\text{VAR}(2)$ ,  $\Phi(L)z_t = (I - \Phi_1 L - \Phi_2 L^2)z_t = \varepsilon_t$  with the following numerical values*

$$\Phi(L) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -0.1 & -0.21 \\ 0.7 & 0.76 \end{bmatrix} L - \begin{bmatrix} 0.3 & 0.8 \\ 0.3 & -0.8 \end{bmatrix} L^2, \quad (5)$$

*the determinant of which is  $-0.48L^4 - 0.189L^3 + 0.571L^2 - 0.66L + 1$  and the adjoint matrix such that*

$$\Phi(L)^{adj} = \begin{bmatrix} 1 - 0.76L + 0.8L^2 & -0.21L + 0.8L^2 \\ 0.7L + 0.3L^2 & 1 + 0.1L - 0.3L^2 \end{bmatrix}.$$

*This involves that the two implied individual series are  $\text{ARMA}(4,2)$ .*

Wallis (1977) has proposed two explanations to solving this paradox. The first solution is to unravel and get rid of common roots in  $\det[\Phi(L)]$  and  $\Phi(L)^{adj}$ . This is not trivial to implement and this is mainly a technical trick for solving the paradox. The second intuition raised by Wallis is based on the observation that coefficients close to zero in the implied ARMA might give the feeling that models are more parsimonious than they theoretically should be. This explanation is always inherent to most of the papers by Zellner and Palm where economic theories are tested and restrictions are imposed on the system which are compatible with implied individual ARMA models. The next subsection provides a new explanation for solving that paradox thanks to the use of short-run co-movements and reduced rank models.

## 2.2 Our contribution to solving the paradox

We introduce a new framework aimed at understanding the gap between implied individual ARMA models deduced from VARs (or VARMA) and the estimated ARIMA we get in empirical studies. More specifically,

---

<sup>2</sup>This result generalizes to individual  $\text{ARMA}(np, (n-1)p + q)$  for  $\text{VARMA}(p, q)$  processes (see e.g. Lütkepohl (2005)).

we investigate the implications of the presence of common cyclical features in a multivariate dynamic model for the order of the marginal processes of the individual series.

The strongest form of common cyclical features is the notion of serial correlation common feature (henceforth, SCCF) proposed by Engle and Kozicki (1993) and Vahid and Engle (1993).<sup>3</sup> In this framework, series  $z_t$  have  $s$  SCCF relationships if there exists an  $n \times s$  matrix  $\delta$  with full column rank and such that  $\delta' \Phi(L) z_t = \delta' z_t = \delta' \varepsilon_t$  in (2). Hence, SCCF implies that impulse response functions of series in  $z_t$  are collinear. Although convenient in terms of parsimony and economic interpretation of business cycles (cycles will be synchronous in the presence SCCF), assumptions underlying SCCF may be too strong. Indeed SCCF implies that the matrices  $\Phi_1, \Phi_2, \dots, \Phi_p$  of  $I - \Phi(L)$  have a common left null space, a condition that has been relaxed *inter alia* in Ahn and Reinsel (1988), Tiao and Tsay (1989), Vahid and Engle (1997), Schleicher (2007), Hecq, Palm and Urbain (2006), Cubadda and Hecq (2001). These latter authors for instance assume that there exists an  $n \times s$  polynomial matrix such that  $\delta'(L) z_t \equiv (\delta_0 + \delta_1 L)' z_t = \delta_0' \varepsilon_t$ . This model is a polynomial SCCF of order one or PSCCF(1), a model that can be generalized (see Cubadda and Hecq (2001) for details) for higher delays of adjustment, a model denoted PSCCF( $m$ ).

For simplicity reason, we start with the most parsimonious model, in which there exists a SCCF matrix. Note again that in the presence of SCCF, the cyclical part of several time series is perfectly synchronized, that is to say that the ups and downs of different economies occur at about the same time. We first illustrate the consequences of the presence of one SCCF relationship in a bivariate example with  $p = 2$  before generalizing to the  $n$  dimensional case for any polynomial degree  $p$ .

**Example 2** *Let us consider the polynomial matrix  $\Phi(L)$  of the Example 1 but now there exists a SCCF relationship with a cofeature vector  $\delta' = (1 : 1)$  for*

$$\Phi(L) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -0.1 & -0.2 \\ 0.1 & 0.2 \end{bmatrix} L - \begin{bmatrix} 0.2 & 0.6 \\ -0.2 & -0.6 \end{bmatrix} L^2. \quad (6)$$

*Because the second and the third coefficient matrices  $\Phi_1$  and  $\Phi_2$  are of reduced rank with the same left null space (second rows in both matrices are the opposite of the first one),  $\delta' \Phi(L) = \delta'$ . The determinant of  $\Phi(L)$  in (6) is  $0.4L^2 - 0.1L + 1$  and the adjoint matrix is*

$$\Phi(L)^{adj} = \begin{bmatrix} -0.2L + 0.6L^2 + 1 & -0.2L + 0.6L^2 \\ 0.1L - 0.2L^2 & 0.1L - 0.2L^2 + 1 \end{bmatrix}.$$

*This implies that the two series are **at most** ARMA(2,2) and not ARMA(4,2) as in the system in Example 1.*

More generally Table 1 summarizes the reduction of the individual ARMA degrees obtained due to common feature restrictions on short-run dynamics. Basically Table 1 reports that multivariate systems

---

<sup>3</sup>This model is also known as a scalar component model of order zero or SCM(0,0) using the terminology of Tiao and Tsay (1989) or white noise direction codependence by Gouriéroux and Peaucelle (1989).

Table 1: Maximum ARMA orders of univariate series generated by a stationary VAR(p) with cofeature restrictions

Models	AR order	MA order
VAR( $p$ )	$np$	$(n-1)p$
SCCF	$(n-s)p$	$(n-s)p$
PSCCF(1)	$(n-s)p + s$	$(n-s)p + (s-1)$
PSCCF( $m$ )	$(n-s)p + sm$	$(n-s)p + (s-1)m$

with additional commonality deliver low order and therefore empirically more plausible, parsimonious ARMA models than unrestricted VAR models. For instance, in a small VAR system with four variables and two lags, implied models are at most ARMA(8,6) while they would reduce to ARMA(2,2) in the presence of a unique common cycle, that is, with  $s = 3$ . This application to Quenouille's results, who already stressed the impact of rank deficiency, is one of the diagnostic tool for data-admissibility we want to formalize and emphasize. Let us formally prove the results that are reported in Table 1.

**Proposition 3 Stationary VAR with  $s$  SCCF.** *In an  $n$ -dimensional stationary VAR( $p$ ) with  $s$  SCCF, the individual ARMA processes have: (i) AR orders not larger than  $(n-s)p$ ; (ii) MA orders not larger than  $(n-s)p$ .*

**Proof.** *Let us rewrite equation (2) as follows*

$$Q(L)x_t = e_t, \quad (7)$$

where  $x_t = Mz_t$ ,  $e_t = M\varepsilon_t$ ,  $Q(L) = M\Phi(L)M^{-1}$ ,  $M' \equiv [\delta, \delta_\perp]$ , and  $\delta_\perp$  is the orthogonal complement. Given that  $x_t$  is a non-singular linear transformation of  $z_t$ , the maximum AR and MA orders of the univariate representation of elements of  $z_t$  must be the same as those of elements of  $x_t$ . Since  $M^{-1} = [\bar{\delta}, \bar{\delta}_\perp]$ , where  $\bar{\delta} = \delta(\delta'\delta)^{-1}$ , and  $\bar{\delta}_\perp = \delta_\perp(\delta'_\perp\delta_\perp)^{-1}$ , we have

$$Q(L) = \begin{bmatrix} I_s & 0 \\ \delta'_\perp\Phi(L)\bar{\delta} & \delta'_\perp\Phi(L)\bar{\delta}_\perp \end{bmatrix}, \quad (8)$$

from which it easily follows that  $\det[Q(L)] = \det[\delta'_\perp\Phi(L)\bar{\delta}_\perp]$  is a polynomial of order  $(n-s)p$ . Hence, the univariate AR order of each element of  $z_t$  is, at most,  $(n-s)p$ . To prove (ii), let  $P(L)$  denote a submatrix of  $Q(L)$  that is formed by deleting one of the first  $s$  rows and one of the first  $s$  columns of  $Q(L)$ . We can partition  $P(L)$  as follows

$$P(L) = \begin{bmatrix} P_{11} & P_{12} \\ P_{21}(L) & P_{22}(L) \end{bmatrix}. \quad (9)$$

Now,  $P_{11}$  is a  $(s-1) \times (s-1)$  identity matrix,  $P_{12}$  is a  $(s-1) \times (n-s)$  matrix of zeros,  $P_{21}(L)$  is a  $(n-s) \times (s-1)$  polynomial matrix of order  $p$ , and  $P_{22}(L)$  is a  $(n-s) \times (n-s)$  polynomial matrix of order  $p$ . Hence,  $\det[P(L)] = \det[P_{11}]\det[P_{22}(L)]$ , which tells us that  $\det[P(L)]$  is of order  $(n-s)p$ . Since cofactors



associated with the blocks of  $Q(L)$  different from  $P_{11}$  are polynomials of degree not larger than  $(n-s)p$ , we conclude that the univariate MA order of each element of  $z_t$  is, at most,  $(n-s)p$ . ■

**Proposition 4 Stationary VAR with  $s$  PSCCF(1).** In an  $n$  stationary VAR( $p$ ) with  $s$  PSCCF(1) we have that the individual ARMA processes have: (i) AR orders not larger than  $(n-s)p + s$ ; (ii) MA orders not larger than  $(n-s)p + (s-1)$ .

**Proof.** Along the same lines of the previous proof, let us consider  $R(L) = M_0 \Phi(L) M_0^{-1}$ ,  $M'_0 \equiv [\delta_0, \delta_{0\perp}]$ , and  $M_0^{-1} \equiv [\bar{\delta}_0, \bar{\delta}_{0\perp}]$ . Then, we have

$$R(L) = \begin{bmatrix} \delta'(L)\bar{\delta}_0 & \delta'_1\bar{\delta}_{0\perp}L \\ \delta'_{0\perp}\Phi(L)\bar{\delta}_0 & \delta'_{0\perp}\Phi(L)\bar{\delta}_{0\perp} \end{bmatrix}.$$

Hence,

$$\det[R(L)] = \det[\delta'(L)\bar{\delta}_0] \det[\delta'_{0\perp}\Phi(L)\bar{\delta}_{0\perp}] - \delta'_{0\perp}\Phi(L)\bar{\delta}_0(\delta'(L)\bar{\delta}_0)^{-1}\delta'_1\bar{\delta}_{0\perp}L.$$

Writing  $(\delta'(L)\bar{\delta}_0)^{-1} = (\delta'(L)\bar{\delta}_0)^{adj} / \det[\delta'(L)\bar{\delta}_0]$  and substituting it above we get

$$\det[R(L)] \underbrace{\det[\delta'(L)\bar{\delta}_0]^{n-s-1}}_{s(n-s-1)} = \det \left\{ \underbrace{\delta'_{0\perp}\Phi(L)\bar{\delta}_{0\perp}}_p \underbrace{\det[\delta'(L)\bar{\delta}_0]}_s - \underbrace{\delta'_{0\perp}\Phi(L)\bar{\delta}_0}_p \underbrace{(\delta'(L)\bar{\delta}_0)^{adj}}_{s-1} \underbrace{\delta'_1\bar{\delta}_{0\perp}L}_1 \right\}$$

from which follows that the degree of the polynomial of  $\det[R(L)]$  is, at most, equal to  $(n-s)p + s$ . Regarding the order of the MA component, let us denote  $G(L)$  a submatrix of  $R(L)$  that is formed by deleting one of the first  $s$  rows and one of the first  $s$  columns of  $R(L)$ . We can partition  $G(L)$  as follows

$$G(L) = \begin{bmatrix} G_{11}(L) & G_{12}L \\ G_{21}(L) & G_{22}(L) \end{bmatrix}, \quad (10)$$

where  $G_{11}(L)$  is a  $(s-1) \times (s-1)$  polynomial matrix of order 1,  $G_{12}$  is a  $(s-1) \times (n-s)$  matrix,  $G_{21}(L)$  is a  $(n-s) \times (s-1)$  polynomial matrix of order  $p$ , and  $G_{22}(L)$  is a  $(n-s) \times (n-s)$  polynomial matrix of order  $p$ . Hence, following a similar reasoning as above, we conclude that the individual MA order is, at most,  $(n-s)p + (s-1)$ . ■

These results can be easily generalized for the PSCCF( $m$ ) case as reported in Table 1. Also note that we do not consider a mixed model that can have both SCCF and PSCCF relationships but results can be trivially deduced from results beneath.

The above propositions can be extended to the case of a I(1) VAR. Let us consider the Vector Error Correction Model (VECM) representation of variables  $z_t$

$$\Gamma(L)\Delta z_t = \alpha\beta' z_{t-1} + \varepsilon_t, \quad (11)$$

where  $\Delta = (1-L)$ ,  $\Gamma(L) = I_n - \sum_{i=1}^{p-1} \Gamma_i L^i$ ,  $\Gamma_i = -\sum_{j=i+1}^p \Phi_j$  for  $i = 1, 2, \dots, p-1$ .  $\alpha$  and  $\beta$  are full-column rank  $n \times r$ -matrices such that  $\Phi(1) = -\alpha\beta'$  and  $\alpha'_\perp \Gamma(1)\beta_\perp$  has full rank. The process  $z_t$  is cointegrated of

order (1,1), denoted by CI(1,1), the columns of  $\beta$  span the cointegrating space, the elements of  $\alpha$  are the corresponding adjustment coefficients, see e.g. Johansen (1996).

A well known implication of cointegration is that  $(n - r - 1)$  unit roots are common to the individual AR and the MA polynomial. Let us formalize this result in order to later extend it to different forms of common cyclical features as well. To do so, consider again equation (2) and apply the transformation

$$B(L)y_t = u_t, \quad (12)$$

where  $y_t = Pz_t$ ,  $u_t = P\varepsilon_t$ ,  $B(L) = P\Phi(L)P^{-1}$ , and  $P' \equiv [\beta, \beta_\perp]$ . Again, the maximum AR and MA orders of the univariate representation of elements of  $z_t$  are the same as those of elements of  $y_t$ . Since  $\beta'z_t$  is I(0) and  $\beta'_\perp z_t$  is I(1), we can partition  $B(L)$  as follows

$$B(L) = \begin{bmatrix} \underbrace{B_{11}(L)}_{r \times r} & \underbrace{\Delta B_{12}(L)}_{r \times (n-r)} \\ \underbrace{B_{21}(L)}_{(n-r) \times r} & \underbrace{\Delta B_{22}(L)}_{(n-r) \times (n-r)} \end{bmatrix}$$

We have that

$$\det[B(L)] = \Delta^{n-r} \det[B_{22}(L)] \det[B_{11}(L) - B_{12}(L)B_{22}(L)^{-1}B_{21}(L)] \equiv \Delta^{n-r} \det[B_{22}(L)] \det[B^*(L)]$$

Since  $B_{22}(1) \neq 0$ , and  $B^*(1) = B_{11}(1) \neq 0$ , there are  $(n-r)$  the unit roots in  $\det[B(L)]$ . However, since  $\beta'_\perp z_t$  is I(1), there must be  $(n-r-1)$  unit roots in the univariate MA polynomials of elements of  $\beta_\perp z_t$ . Hence, the AR and MA orders of elements of  $z_t$  respectively are, at most,  $n(p-1) + r + 1$  and  $(n-1)(p-1) + r$ .

For the VECM in (11), the presence of  $s$  SCCF's implies that there exists an  $n \times s$  full-rank matrix  $\delta$  such that  $\delta' \Delta z_t = \delta' \varepsilon_t$  or equivalently

$$\delta' \Phi(L) = \delta' \Delta.$$

The existence of  $s$  PSCCF(1) relationships requires that there exists an  $n \times s$  full-rank matrix  $\delta_0$  such that  $\delta'_0 \Delta z_t + \delta'_1 \Delta z_{t-1} = \delta'_0 \varepsilon_t$  where  $\delta_1 = \Gamma'_1 \delta_0$  or equivalently

$$\delta(L)' \Phi(L) = \delta'_0 \Delta.$$

Finally, the presence of  $s$  weak form SCCF (WF's) implies that here exists an  $n \times s$  full-rank matrix  $\tilde{\delta}$  such that  $\tilde{\delta}' \Delta z_t - \tilde{\delta}' \alpha \beta' z_{t-1} = \tilde{\delta}' \varepsilon_t$  or equivalently

$$\tilde{\delta}(L)' \Phi(L) = \tilde{\delta}',$$

where  $\tilde{\delta}(L) = \tilde{\delta} + \tilde{\delta}_1 L$ ,  $\tilde{\delta}_1 = -(\beta \alpha' + I_n) \tilde{\delta}$  (see Cubadda, 2007).

Table 2 reports the largest ARIMA orders of individual series that are generated by a VECM with common feature restrictions. Let us formally prove these results.

Table 2: Maximum ARIMA orders of univariate series generated by a CI(1,1) VAR(p) with cofeature restrictions

Models	AR order	Integration order	MA order
VAR(p)	$n(p-1) + r$	1	$(n-1)(p-1) + r$
SCCF	$(n-s)(p-1) + r$	1	$(n-s)(p-1) + r$
PSCCF(1)	$(n-s)(p-1) + r + s$	1	$(n-s)(p-1) + r + s - 1$
WF	$(n-s)(p-1) + r$	1	$(n-s)(p-1) + r$

**Proposition 5 Cointegrated VAR with s SCCF.** In a CI(1,1) VAR(p) with s SCCF we have that the individual ARMA processes of the first differences have: (i) AR orders not larger than  $n(p-1) + r$ ; (ii) MA orders not larger than  $(n-1)(p-1) + r$ .

**Proof.** By construction of  $M$ , and keeping in mind that  $\delta' \Phi(L) = \delta' \Delta$ , we notice that

$$\det[Q(L)] = \det \begin{bmatrix} \Delta I_s & 0 \\ \delta'_\perp \Phi(L) \bar{\delta} & \delta'_\perp \Phi(L) \bar{\delta}_\perp \end{bmatrix} = \Delta^s \det[\delta'_\perp \Phi(L) \bar{\delta}_\perp]$$

is a polynomial of order  $(n-s)p + s$ . Since  $\det[Q(L)]$  is a polynomial of the same order as  $\det[\Phi(L)]$ , we conclude that the univariate AR order of each element of  $z_t$  is, at most,  $(n-s)p + s$ . Note that this result means that the AR part is of order  $(n-s)(p-1) + n$ . The polynomial degree of the moving average component is obtained along the same line as in Proposition 3 but with  $P_{11}\Delta$  instead of  $P_{11}$  in (9). Hence,  $\det[P(L)] = \det[P_{11}\Delta] \det[P_{22}(L)]$ , which tells us that  $\det[P(L)]$  is of order  $(n-s)p + (s-1)$ . Since the presence of cointegration implies that  $(n-r-1)$  common unit roots cancel out from the individual AR and the MA polynomials, the result follows. ■

**Proposition 6 Cointegrated VAR with s PSCCF(1).** In a CI(1,1) VAR(p) with s PSCCF(1) we have that the individual ARMA processes of the first differences have: (i) AR orders not larger than  $(n-s)(p-1) + s + r$ ; (ii) MA orders not larger than  $(n-s)(p-1) + r + s - 1$ .

**Proof.** Since it is similar to that of Proposition 4, it is not reported to save space. ■

**Proposition 7 Cointegrated VAR with s WF.** In a CI(1,1) VAR(p) with s WF we have that the ARMA processes of the first differences have: (i) AR orders not larger than  $n(p-1) + r$ ; (ii) MA order not larger than  $(n-1)(p-1) + r$ .

**Proof.** Let us consider  $H(L) = \widetilde{M} \Phi(L) \widetilde{M}^{-1}$ ,  $\widetilde{M}' \equiv [\widetilde{\delta}, \widetilde{\delta}_\perp]$ , and

$$H(L) = \begin{bmatrix} \widetilde{\delta}'(L) \widetilde{\delta} (\widetilde{\delta}' \widetilde{\delta})^{-1} & \widetilde{\delta}'_1 \widetilde{\delta}_\perp (\widetilde{\delta}'_\perp \widetilde{\delta}_\perp)^{-1} L \\ \widetilde{\delta}'_\perp \Phi(L) \widetilde{\delta} (\widetilde{\delta}' \widetilde{\delta})^{-1} & \widetilde{\delta}'_\perp \Phi(L) \widetilde{\delta}_\perp (\widetilde{\delta}'_\perp \widetilde{\delta}_\perp)^{-1} \end{bmatrix}$$

Hence, the proof is entirely analogous to that of Proposition 4 with the  $H(L)$  matrix in place of  $R(L)$ . ■

The next subsection generalizes some of these results for blocks of countries.

### 2.3 Corollaries for block-diagonal VARs and the separation hypothesis

From the discussion above, naturally arises the question about block-diagonal systems, namely about a situation in which we can disentangle groups of variables<sup>4</sup> having many within co-movements but that are roughly independent to other sets of variables. We make the distinction between two types of models we illustrate now for SCCF restrictions. For the sake of simplicity, we consider two groups of variables.

The first representation is a block-diagonal stationary VAR( $p$ ) with SCCF within each block of variables  $n_1$  and  $n_2$  respectively, with  $n_1 + n_2 = n$  such that

$$\Phi(L) = \begin{bmatrix} \Phi_{11}(L) & 0 \\ 0 & \Phi_{22}(L) \end{bmatrix}, \quad (13)$$

where there exist two full column rank matrices with ranks  $s_1$  and  $s_2$  respectively, with  $s_1 + s_2 = s$  such that  $\delta'_1 \Phi_{11}(L) = \delta'_1$  and  $\delta'_2 \Phi_{22}(L) = \delta'_2$ . This is a system with separability in common features (Hecq, Palm and Urbain, 2002) with a block-diagonal co-feature matrix such that

$$\underbrace{\delta}_{n \times s} = \begin{bmatrix} \underbrace{\delta_1}_{n_1 \times s_1} & \underbrace{0}_{n_1 \times s_2} \\ \underbrace{0}_{n_2 \times s_1} & \underbrace{\delta_2}_{n_2 \times s_2} \end{bmatrix}.$$

The co-feature matrices  $\delta_i$  could be but need not to be identical. From Table 1 it is known that the maximum degrees for the implied univariate ARMA processes are  $(n_1 + n_2 - s_1 - s_2)p$  for both AR and MA components, assuming the same polynomial order  $p$  for each group. With different polynomial orders  $p_1$  and  $p_2$ , it is  $(n_1 - s_1)p_1 + (n_2 - s_2)p_2$  respectively. However it is clear that these degrees can be further reduced due to the presence of blocks of zeros. Because this leads to the presence of common roots between the AR and MA parts, the maximum orders of both AR and MA components are  $(n_1 - s_1)p_1$  for the first block and  $(n_2 - s_2)p_2$  for the second group of series.

Due to separation in the form of block-diagonality of  $\Phi(L)$ , the implied univariate ARMA-processes need not to have identical AR-polynomials. In fact, the model allows for inter-block AR parameter heterogeneity and intra-block AR parameter homogeneity. Note that this form of separation is often used when studying data for a set of countries using a panel-data framework. In some panel-data studies, when a single variable is analyzed for a set of countries which can be clustered in groups with intergroup block-diagonality of  $\Phi(L)$  combined with intragroup homogeneity so that the univariate ARMA models for a given group have identical AR polynomials.

The second model is a triangular VAR such that

$$\Phi(L) = \begin{bmatrix} \Phi_{11}(L) & \Phi_{12}(L) \\ 0 & \Phi_{22}(L) \end{bmatrix}, \quad (14)$$

---

<sup>4</sup>They could for instance correspond to a group of countries.

with  $\delta'_1[\Phi_{11}(L) : \Phi_{12}(L)] = \delta'_1$  and  $\delta'_2\Phi_{22}(L) = \delta'_2$ . The situation is similar as before because  $\delta_\perp$  is still block-diagonal. However fewer cancellations of common roots are observed and the implied models are  $(n_1 - s_1)p_1 + (n_2 - s_2)p_2$  for block 1 and  $(n_2 - s_2)p_2$  in the block 2.

In this case, the AR polynomials of the two blocks could have factors in common which are identical to the AR polynomial of the second block. Notice also that under the additional assumption of appropriate block-diagonality of the contemporaneous covariance matrix  $\Sigma$ , the  $n_2 \times 1$  subvector  $x_t$  of  $z_t = (y'_t, x'_t)'$  is strongly exogenous. Multiplying the subsystem for  $y_t$  by the adjoint matrix of  $\Phi_{11}(L)$ , one obtains the set of transfer functions (TF) for  $y_t$ , that is a set of dynamic equations with the same scalar AR polynomial, and vector moving averages in  $x_t$  and in  $\varepsilon_{1t}$ , with  $\varepsilon_{it}$  being the subvector of  $\varepsilon_t$  corresponding to the partitioning of  $z_t$  as  $(y'_t, x'_t)'$ . These TF-equations can be used to study the dynamics of  $y_t$  or each of its components given  $x_t$  and its past and the past of  $y_t$ .

Obviously, similar results can be obtained for the general case of block-diagonal or block-triangular systems with  $k$  blocks.

## 2.4 Interpretation in terms of a VMA index model

This subsection further investigates the consequences of the presence of short-run co-movements for the VMA part of final equations. We consequently look at the adjoint matrix of  $\Phi(L)$  and emphasize that the VMA representation has a particular form and follows a sort of multivariate index model (see Reinsel, 1983). In order to introduce the problem, let us look more closely at the adjoint matrix  $\Phi(L)^{adj}$ , we have computed in Example 2 for  $s = 1$  SCCF relationship, i.e.

$$\Phi(L)^{adj}\varepsilon_t = \begin{bmatrix} 1 - 0.2L + 0.6L^2 & -.2L + .6L^2 \\ 0.1L - 0.2L^2 & 1 + 0.1L - 0.2L^2 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

We can also write the previous expression in terms of coefficient matrices and we observe the presence of a common right factor, namely of common right null spaces such that

$$\Phi(L)^{adj}\varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} + \begin{bmatrix} -.2 \\ 0.1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{bmatrix} + \begin{bmatrix} 0.6 \\ -0.2 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-2} \\ \varepsilon_{2t-2} \end{bmatrix}$$

with the obvious numerical observation that  $\Phi(L)^{adj}\delta_\perp = \delta_\perp$ ,  $\delta'_\perp = (1 : -1)$ . We show now that the VMA component of (3) has a structure that is analogous to the multivariate index model of Reinsel (1983), and this leads to the following proposition of which this example is only a particular case.

**Proposition 8** *In a stationary VAR(p), the existence of s SCCF vectors implies an index model in the MA matrices of polynomial degree strictly larger than  $(n - s - 1)p$  with a common right null space generated by  $\delta$ . This implies that post-multiplying the adjoint matrix of  $\Phi(L)$  by  $\delta_\perp$  reduces the order of the MA part to a degree of at most  $(n - s - 1)p$  instead of  $(n - s)p$ .*

**Proof.** We denote  $Q(L) = M'\Phi(L)M$  with  $M \equiv [\delta : \delta_\perp]$

$$Q(L) = \begin{bmatrix} \delta'\delta & 0 \\ \delta'_\perp\Phi(L)\delta & \delta'_\perp\Phi(L)\delta_\perp \end{bmatrix}.$$

Let us compute  $Q(L)^{-1}$  using the definition of partitioned matrices, namely

$$Q(L)^{-1} = \begin{bmatrix} (\delta'\delta)^{-1} & 0 \\ -(\delta'_\perp\Phi(L)\delta_\perp)^{-1}(\delta'_\perp\Phi(L)\delta)(\delta'\delta)^{-1} & (\delta'_\perp\Phi(L)\delta_\perp)^{-1} \end{bmatrix}.$$

Because  $Q(L)^{-1} = M^{-1}\Phi(L)^{-1}M^{-1'}$  and we know from orthogonal spaces of  $M$  that  $M^{-1} = [(\delta'\delta)^{-1}\delta' : (\delta'_\perp\delta_\perp)^{-1}\delta'_\perp]$  we obtain the inverse of  $\Phi(L)$  using  $\Phi(L)^{-1} = MQ(L)^{-1}M'$ , namely

$$\begin{aligned} \Phi(L)^{-1} &= \delta(\delta'\delta)^{-1}\delta' - \delta_\perp(\delta'_\perp\Phi(L)\delta_\perp)^{-1}(\delta'_\perp\Phi(L)\delta)(\delta'\delta)^{-1}\delta' \\ &\quad + \delta_\perp(\delta'_\perp\Phi(L)\delta_\perp)^{-1}\delta'_\perp. \end{aligned}$$

The adjoint matrix is simply  $\Phi(L)^{-1}\det[\Phi(L)] = \Phi(L)^{adj}$ . Post-multiplying by  $\delta_\perp$  gives

$$\begin{aligned} \Phi(L)^{adj}\delta_\perp &= \det[\Phi(L)]\delta_\perp(\delta'_\perp\Phi(L)\delta_\perp)^{-1}\delta'_\perp\delta_\perp = \\ &= \det[\Phi(L)]\delta_\perp \frac{(\delta'_\perp\Phi(L)\delta_\perp)^{adj}}{\det[\delta'_\perp\Phi(L)\delta_\perp]}\delta'_\perp\delta_\perp \\ &= \det[\Phi(L)]\delta_\perp \frac{(\delta'_\perp\Phi(L)\delta_\perp)^{adj}}{\det[\delta'_\perp\delta_\perp(I_{n-s} - \Psi'(L)L\delta_\perp)]}\delta'_\perp\delta_\perp \\ &= \det[\Phi(L)]\delta_\perp \frac{(\delta'_\perp\Phi(L)\delta_\perp)^{adj}}{\det[\delta'_\perp\delta_\perp]\det[(I_{n-s} - \Psi'(L)L\delta_\perp)]}\delta'_\perp\delta_\perp \\ &= \delta_\perp \frac{(\delta'_\perp\Phi(L)\delta_\perp)^{adj}}{\det[\delta'_\perp\delta_\perp]}\delta'_\perp\delta_\perp, \end{aligned}$$

where  $\Phi(L) = [I_n - \delta_\perp\Psi'(L)L]$  the degree of the polynomial in  $(\delta'_\perp\Phi(L)\delta_\perp)^{adj}$  is  $(n-s-1)p$ . ■

**Corollary 9** In the particular case with  $n-1 = s$ , this combination yields a white noise.

The last result is more a mathematical curiosity than a device to be used in an empirical analysis. We might think using it if one wants fully efficient estimation of the VARMA for instance. But this shows that there exists a factor representation in the MA part.

In empirical investigations, this result can partially explain why the MA part is smaller than what it could be, given theoretical implied models in Tables 1 and 2. An interesting case arises when shocks are almost singular. Indeed, for instance consider the VMA(1)  $w_t = \varepsilon_t + \Phi_1^{adj}\varepsilon_{t-1}$  denoted  $w_t = \varepsilon_t + H_1\varepsilon_{t-1}$  to make notation easier. We have the first autocovariance given by  $E(w_tw'_{t-1}) = H_1\Sigma$ . Because  $\delta$  spans the right null space of  $H_1$ ,  $E(w_tw'_{t-1}) \approx 0$  in case of near perfect collinearity of  $\varepsilon_t$  namely when  $\Sigma = \delta_\perp K$ . The value of the correlation depends on the sign of the cofeature relationship that forms the combination. For instance, when there exists a negative relationship between the first differences of both hypothetical series,

$\delta' = (1 : 1)$  and consequently we also have a condition on the disturbances via  $\delta_\perp$ . In the application with the US and Canada, we have  $\hat{\delta}' \approx (1 : -1)$  and a positive correlation between the residuals of the VAR (around 0.65), this is why the MA(1) was close to being negligible. Of course the relationship between variables is never collinear, this is why we call it a near coincidental situation.

**Remark 10** *Note that the expression of the adjoint can be simplified as follows. Let us use*

$$\begin{aligned}\Phi(L)^{adj} &= \det[\Phi(L)]\delta(\delta'\delta)^{-1}\delta' \\ &\quad - \det[\Phi(L)]\delta_\perp(\delta'_\perp\Phi(L)\delta_\perp)^{-1}(\delta'_\perp\Phi(L)\delta)(\delta'\delta)^{-1}\delta' \\ &\quad + \det[\Phi(L)]\delta_\perp(\delta'_\perp\Phi(L)\delta_\perp)^{-1}\delta'_\perp.\end{aligned}$$

*The second line is equal zero and we have*

$$\begin{aligned}\Phi(L)^{adj} &= \det[\Phi(L)]\delta(\delta'\delta)^{-1}\delta' \\ &\quad + \det[\Phi(L)]\delta_\perp(\delta'_\perp\delta_\perp)^{-1}\delta'_\perp\Phi(L)^{-1}.\end{aligned}$$

*To obtain these results we use the properties that  $(\delta'_\perp\Phi(L)\delta_\perp)^{-1} = (\delta'_\perp\delta_\perp)^{-1}(I_{n-s} - \Psi'(L)L\delta_\perp)^{-1}$  or to simplify notations  $(\delta'_\perp\delta_\perp)^{-1}\Xi(L)^{-1}$ . We also use the property  $(\delta'_\perp\delta_\perp)^{-1}\Xi(L)^{-1}\delta'_\perp = (\delta'_\perp\delta_\perp)^{-1}\delta_\perp\Phi(L)^{-1}$  (see inter alia Lütkepohl (1996, p. 29)). To provide a simple check, simply premultiply the adjoint by  $\Phi(L)$  to see that  $\det[\Phi(L)] \equiv \det[\Phi(L)][\delta(\delta'\delta)^{-1}\delta' + \delta_\perp(\delta'_\perp\delta_\perp)^{-1}\delta'_\perp]$ .*

## 2.5 An example: Industrial output growth in Canada and the US

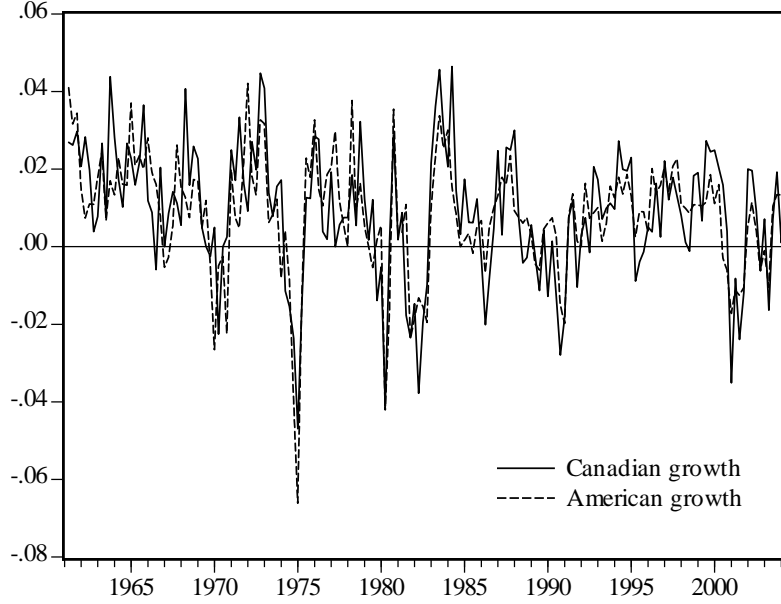
Engle and Kozicki (1993) find that there exists a SCCF relationship between the Canadian and the US quarterly growth rates of output. They have considered a sample from the late 1950s to the late 1990s. We select the same countries and we use the seasonally adjusted industrial production indexes from OECD main indicators over the period 1960:Q1-2004:Q3, namely we have 175 observations. Figure 1 plots these series. The model selection criteria LR, AIC, HQ and SBC lead to selecting  $p = 2$  for the log-levels of the bivariate processes. We reject the presence of cointegration at usual significance levels using Johansen's trace test. Consequently the analysis is performed in first differences, namely with quarterly growth rates, with one lag only ( $p - 1 = 1$ ).

The estimation by OLS of the VAR(1) in first differences delivers (standard errors in brackets)

$$\begin{bmatrix} \Delta \ln \widehat{USA}_t \\ \Delta \ln \widehat{CA}_t \end{bmatrix} = \begin{bmatrix} 0.003 \\ (0.001) \\ 0.003 \\ (0.001) \end{bmatrix} + \begin{bmatrix} 0.333 & 0.273 \\ (0.088) & (0.079) \\ 0.265 & 0.360 \\ (0.102) & (0.092) \end{bmatrix} \begin{bmatrix} \Delta \ln USA_{t-1} \\ \Delta \ln CA_{t-1} \end{bmatrix}.$$

In order to check whether the first autoregressive matrix is of reduced rank we compute a SCCF test statistics using a canonical correlations approach (e.g. Vahid and Engle, 1993) between  $\Delta z_t = (\Delta \ln USA_t : \Delta \ln CA_t)'$  and  $\Delta z_{t-1}$ . The results are as follows:  $p - value = 0.31$  is the p-value associated with the null hypothesis ( $\chi^2_{(1)}$ ) that a linear combination of  $\Delta z_t$  is orthogonal to the past of  $\Delta z_t$ . Information criteria also lead to

Figure 1: Quarterly growth rates of industrial production indexes (industrial sector)



select  $s = 1$  (Hecq, 2006). The estimated common cyclical feature relationship is  $\Delta \ln USA_t - 1.05 \Delta \ln CA_t$ . Without any constraints on the short-run, both implied series should follow ARMA(2,1) processes. However, the sample ACF and PACF indicate (figures not reported in the paper) that the orders are probably shorter and that AR(1) models for industrial production in both countries are more appropriate. For univariate AR(1) models, the estimated equations have quite similar autoregressive roots

$$\begin{aligned} \Delta \ln \widehat{USA}_t &= \frac{0.003}{(0.001)} + \frac{0.554}{(0.062)} \Delta \ln USA_{t-1}, \\ \Delta \ln \widehat{CA}_t &= \frac{0.004}{(0.001)} + \frac{0.533}{(0.064)} \Delta \ln CA_{t-1}, \end{aligned}$$

where for both equations, the null of no disturbance autocorrelation is not rejected using LM tests. Note that although parameters are consistently estimated by OLS (or by single equation ML in case there is an MA component), the standard  $t$ -ratios we report in brackets do not lead to the most powerful test because errors are serially cross-correlated and OLS is not efficient.

One part of our story is confirmed, that is to say the implied individual ARMA models with additional commonality in the dynamics have a shorter autoregressive part then we should get without these restrictions. Moreover the autoregressive roots are similar for both countries. What might look as evidence against our prediction is the absence of a MA part because both growth rates should be ARMA(1,1).<sup>5</sup> First note that

---

<sup>5</sup>Using univariate maximum likelihood estimation, the MA(1) component has p-values of 0.16 and 0.32 for respectively the



these numbers are maximum degree (the meaning of at most) and second that empirical shorter models may be found due to specific coefficient values. For instance this might be linked to phenomena similar to those presented in Subsection 2.4.

### 3 Estimation procedures and a small sample evaluation

We address the problem of estimation and testing of single univariate ARMA models and sets of univariate ARMA models implied by a VAR( $p$ ) model in the presence of cofeature restrictions. Univariate ARMA models can be estimated in a straightforward way by the maximum likelihood method. Therefore, we identify and estimate by ML for each series individually the parsimonious empirical ARMA( $p_i, q_i$ ) such that

$$\hat{z}_{it} = \hat{\alpha}_i + \sum_{j=1}^{p_i} \hat{\phi}_{ij} z_{it-j} + \sum_{k=1}^{q_i} \hat{\theta}_{ik} \hat{u}_{it-k}, \quad i = 1 \dots n, \quad t = 1 \dots T, \quad (15)$$

where  $\hat{\alpha}_i$ ,  $\hat{\phi}_{ij}$  and  $\hat{\theta}_{ik}$  are estimated scalar parameters for series  $i$ ,  $i \in \{1, 2, \dots, n\}$ ;  $p_i$  and  $q_i$  are the lag orders of the ARMA model for the  $i$ th series and they might empirically differ from series to series. This first stage of the analysis is helpful because we can have a first idea about a maximum autoregressive degree  $p$  as well as of the number of series that might share co-movements.<sup>6</sup> So doing, we obtain a rough indication, a sort of bound say, regarding the possible number of common feature vectors, with  $\max(p_i) = (n - s)p$ . Also, if for a series the autoregressive order differs much from that for the other series, then probably it does not share a common cycle with these other variables. Indeed, in absence of cancelling and other coincidental cases, the  $\phi_{ij}$ 's should be the same for all  $z_{it}$ 's.

For sets of univariate ARMA models derived from a stationary multivariate ARMA model, Wallis (1977) has considered maximum likelihood estimation, whereas Palm and Zellner (1980) have considered both maximum likelihood and efficient two-step estimation. For (difference)-stationary series, the asymptotic properties of univariate and multivariate estimation and test procedures are standard and known. For unit root processes with or without cointegration present, when the unit root is not imposed, the asymptotic procedures generally are non-standard. Whether standard asymptotics hold or not, likelihood and efficient two-step methods for systems may be cumbersome to implement when the set of variables is large. Therefore, for empirical work, it will be useful to have tests at the disposal which can be easily implemented and which lead to reliable inference. Satisfying this need is the objective of this section.

#### 3.1 Estimation of the common autoregressive component and aggregating individual series

It has been emphasized in Section 2 that the series implied by the VAR must have the same autoregressive coefficients. Consequently we further investigate this restriction using an estimator of the common

---

US and Canada.

<sup>6</sup>Note that even though we do not estimate the VAR, and this is in the spirit of what we want to do in practice, we may nonetheless have an idea about  $p$ . For instance with annual data  $p$  is usually at most three or four but  $p = 2$  is more likely to be found; for quarterly data 8 or 10 is a maximum although it is generally less than that in empirical work.

autoregressive part<sup>7</sup> such that

$$\hat{z}_{it} = \hat{\alpha}_i + \sum_{j=1}^{p_i} \hat{\phi}_j z_{it-j} + \sum_{k=1}^{q_i} \hat{\theta}_{ik} \hat{u}_{it-k}, \quad (16)$$

where  $\hat{\phi}_j$  is the  $j$ th lag order common to all  $n$ . Also note that in (16), the estimated intercept and moving average parameters are numerically different from (15) but we leave it like this to avoid introducing new notations. The intuition underlying the use of (16) is twofold. First if we are under the null hypothesis, then this is a correct way to impose the commonality observed in different series. Secondly, under the alternative namely when including a set of  $n_1$  variables that does not belong to the same multivariate process then, erroneously imposing a common autoregressive root for the  $n - n_1$  variables leads to leave some extra autoregressive terms in the error process. As a result, the detected order of the MA part can be higher than those observed in individual components. But before doing this we need to get estimates of the common autoregressive component.

After having identified and estimated the parsimonious empirical ARMA( $p_i, q_i$ ) by ML for each series individually, one solution is to compute the average of  $n$  autoregressive parameters such that  $\hat{\phi}_j^{mg} = n^{-1} \sum_{i=1}^n \hat{\phi}_{ij}$ ,  $j = 1 \dots \max(p_i)$ . However proceeding this way the homogeneity we have under the null is ignored. Moreover the estimation of  $n$  equations by ML induces a lot of variability in the estimation of that average.

Instead, because the implied model for  $\det[\Phi(L)]\bar{Z}_t$  is at most a MA $[(n-s)p]$  model in the presence of SCCF vectors for instance, we further use this observation and estimate an ARMA model for the average of the  $n$  series in the ML estimation of

$$\bar{z}_t = \hat{\alpha} - \sum_{j=1}^{p_i} \hat{\phi}_j^{av} \bar{z}_{t-j} + \sum_{k=1}^{q_i} \hat{\theta}_k \hat{v}_{t-k} + \hat{v}_t, \quad (17)$$

with  $\bar{z}_t = n^{-1} \sum_{i=1}^n z_{it}$  the simple average of  $n$  series. In this second setting we might use a LM test for no autocorrelation or other graphical tools (i.e. ACF, PACF) on  $\hat{v}_t$  to check the white noise hypothesis. A rejection of that null is a sign of misspecification, namely that probably we have included in the analysis a series that is not implied by the same system and consequently does not have the same final equation or does not have features in common with the other series in the system and therefore has higher degree AR and MA polynomials.

Cubadda, Hecq and Palm (2007) have shown in simulations that estimating a parsimonious ARMA model on aggregates is the best strategy for obtaining a common autoregressive coefficient. Indeed, this procedure not only imposes the common AR parameters but also reduces the MA parts that might be annihilated by linearly combining the series (see Section 3). Using the averages for Canada and the US with  $\Delta \bar{z}_t = \frac{1}{2}(\Delta \ln USA_t + \Delta \ln CA_t)$ , the best model (using SBC) for the aggregates is the AR(1)

$$\Delta \bar{z}_t = 0.016 + \underset{(3.55)}{0.617} \Delta \bar{z}_{t-1} + \hat{e}_t, \quad \underset{(10.40)}$$

---

<sup>7</sup>Note that to obtain the common AR operator, people have proposed to factor the AR-polynomials for the individual series into the product of the individual operators and then to cancel common factors (Lütkepohl, 2005, 496).

where  $t$ -ratios are in brackets. LM autocorrelation tests on the residuals do not reject the null at any common significance level.

Moreover, aggregation can also yield an additional "virtuous" property for such models. Indeed we have focused in Subsection 2.4 on the right multiplication of the adjoint by a vector or a matrix. More interesting from an empirical perspective, is the combination on the left, hence cross-sectional aggregation. Let us denote  $\kappa = \frac{1}{n}\iota$  with  $\iota = (1 \ 1 \dots 1)'$  an  $n$  vector of ones,  $\kappa' \det[\Phi(L)]z_t = \det[\Phi(L)]\bar{z}_t$  computes that average. This operation can reduce in some circumstances the degree of the MA component. Actually to further reduce the degree of the MA component we must have  $\kappa \in sp(f_p)$  where in the stationary VAR( $p$ ) under SCCF<sup>8</sup>,  $f_p$  is the matrix

$$z_t = \delta_{\perp} f_1' z_{t-1} + \dots + \delta_{\perp} f_p' z_{t-p} + \varepsilon_t.$$

Intuitively, in a set of  $n$  time series with  $s$  SCCF, there exists a combination formed by the column of  $f_p$  that reduces the dynamics in the MA component from  $(n-s)p$  to  $(n-s)p-1$ . It may be reduced to  $(n-s)p-2$  (and so on) if for the last two  $sp(f_p) = sp(f_{p-1})$ . This comes from the fact that the adjoint matrix must have the form  $\varepsilon_t = f_{1\perp} G_1 \delta' \varepsilon_{t-1} + \dots + f_{p\perp} G_p \delta' \varepsilon_{t-p} + \varepsilon_t$  with  $G_i$  full rank normalizing square matrices. Indeed otherwise the polynomial degree in  $\Phi(L)\Phi(L)^{adj} > \text{degree } \det[\Phi(L)]$  given in Table 1 for instance. Each premultiplication by a  $f_i$  cancels the corresponding order in the adjoint. But to reduce the dynamics to the maximum order we are interested in taking  $f_p$ . When the last two matrices of the VAR have the same right null space, i.e. with  $f_{p-1} = f_p$ , premultiplying the adjoint by  $f_p$ , the lag order will be decreased by 2.

**Example 11** *Let us consider the stationary case with SCCF and with  $n = 2, s = 1, p = 2$  such that the maximal orders for the AR and the MA components are 2 and the normalizing matrices are  $I$ . We have  $\Phi(L)\Phi(L)^{adj} = (I - \delta_{\perp} f_1' L^1 - \delta_{\perp} f_2' L^2)(I - f_{1\perp} \delta' L^1 - f_{2\perp} \delta' L^2)$ . This gives  $I - \delta_{\perp} f_1' L^1 - \delta_{\perp} f_2' L^2 - \delta_{\perp} f_1' L^1 + \delta_{\perp} f_1' f_{2\perp} \delta' L^3 + \delta_{\perp} f_2' f_{1\perp} \delta' L^3 - \delta_{\perp} f_2' L^2$ . Because we have  $\delta_{\perp} f_1' f_{2\perp} \delta' L^3 = -\delta_{\perp} f_2' f_{1\perp} \delta' L^3$ , the degree is indeed at most 2.*

Ideally  $f_p$  should be estimated, say using a canonical correlation procedure for instance. However not only this is not always feasible but our objective is to make the analysis as simple as possible. Consequently we will use the simple average combination  $\kappa$  as a pre-testing tool in empirical work. Accepting the null using  $\kappa$  will be considered as a sign of the presence of co-movements (and then that  $sp(f_p) = \kappa$ ) although rejecting the same hypothesis will not be seen as an absence of such relationships.

Averaging series is fine under the null that the final equations are from the same initial VAR model. Next we evaluate whether this is also true when we "erroneously substitute" to the group of countries having common cycles, additional series from another group.

---

<sup>8</sup>This is also valid for the VECM.

### 3.2 A Monte Carlo experiment on the estimator on aggregates

Let us look at the final equations of a VAR(1) with a reduced rank structure

$$z_t = \alpha + \Phi_1 z_{t-1} + \varepsilon_t = \alpha + \delta_{\perp} C_1' z_{t-1} + \varepsilon_t. \quad (18)$$

In order to keep fixed the AR and the MA degrees in the  $n$  implied models we impose such as in Cubadda, Hecq and Palm (2007) that there exist  $s = n - 1$  SCCF relationships, leading to  $n$  implied ARMA(1,1) models whatever  $n$  instead of ARMA( $np, (n - 1)p$ ). Our associated common feature matrix assumes full short-run convergence between economies (or variables). Indeed, one of the issues we want to emphasize is to determine a core of countries (or variables) with the maximum number of co-movements. The cofeature matrix has thus the shape, for instance with  $n = 5$  :

$$\delta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -1 & -1 & -1 \end{pmatrix}.$$

leading to (up to a normalization)  $\delta_{\perp} = (1 : 1 : 1 : 1 : 1)'$ .  $\alpha$  is a  $n$  dimensional vector of constant terms we generate from a uniform distribution on (0,1). The factor  $C_1$  cannot change at each replication otherwise we could not compute a bias because implied roots would be different at each of these replications. In order to have the same coefficient whatever the odd number of individuals we choose  $C_1' = (0.5, -0.5, 0.5, -0.5, 0.5, \dots -0.5, 0.5)$ . This guarantees both the stationarity of the multivariate process and the common value of the AR root ( $\phi_{i1} = 0.5, i = 1 \dots n$ ) for every series whatever the number of individuals. To show this we can simply use the property of partitioned matrices and compute  $\det[I - \Phi_1 L]$  when considering the  $(n - 1) \times (n - 1)$  upper-left block of  $(I - \Phi_1 L)$ . For the covariance matrix of the VAR we use  $\Sigma_{\varepsilon} = I$ . This implies cross-correlated errors  $w_t = \Phi(L)^{adj} \varepsilon_t$  in the final equations because their contemporaneous covariance matrix is  $E(w_t w_t') = I + \Phi_1 \Phi_1'$ . This corresponds to a correlation between the disturbances of the implied equations of  $\rho = 0.55$  for all pairs. We use  $M = 2000$  replications and generate  $T + 50$  observations by iteration before dropping the first 50 points to initialize the random sequence.

We consider successively  $n = 5, 11, 23$  individuals for  $T = 50, 100, 250$  data points. However we substitute the last 2nd, 5th and 11th series by an heterogenous diagonal VAR generated from

$$y_t = \Gamma_1^j y_t + \varepsilon_t, \quad t = 1 \dots T, \quad j = a, b, c \quad (19)$$

with successively:  $\Gamma_1^a = \text{diag}(0.5) \pm U(-0.125, 0.125)$ ,  $\Gamma_1^b = \text{diag}(0.25) \pm U(-0.125, 0.125)$  and  $\Gamma_1^c = \text{diag}(0) \pm U(-0.125, 0.125)$ . These are heterogeneous AR(1) processes but only  $\Gamma_1^a$  is centered around the same value as those from the final equations, namely 0.5.

Table 3 reports both the empirical bias and RMSE when we combine a system (18) of  $n_1$  equations with a

Table 3: Estimation of a common autoregressive parameter in a mixed system

<i>DGP</i>		<i>Bias</i>			<i>rmse</i>		
$(n_1 + n_2)/T$		50	100	250	50	100	250
3+2	$\Gamma_1^a$	-0.074	-0.034	-0.010	0.196	0.124	0.068
	$\Gamma_1^b$	-0.100	-0.060	-0.036	0.214	0.139	0.078
	$\Gamma_1^c$	-0.122	-0.078	-0.056	0.225	0.152	0.093
6+5	$\Gamma_1^a$	-0.088	-0.048	-0.032	0.207	0.124	0.075
	$\Gamma_1^b$	-0.098	-0.058	-0.042	0.213	0.131	0.082
	$\Gamma_1^c$	-0.106	-0.067	-0.048	0.219	0.137	0.087
12+11	$\Gamma_1^a$	-0.080	-0.041	-0.015	0.207	0.137	0.067
	$\Gamma_1^b$	-0.083	-0.043	-0.018	0.212	0.136	0.068
	$\Gamma_1^c$	-0.085	-0.044	-0.020	0.213	0.137	0.069

Note: A parsimonious empirical ARMA is chosen on the aggregates using SBC. The DGP is such that from the  $n = n_1 + n_2$  final equations we only consider the first  $n_1$  set and we add  $n_2$  heterogeneous AR(1) processes centered around 0.5 (a), 0.25 (b) and 0 (c) in  $\Gamma_1^a, \Gamma_1^b, \Gamma_1^c$ .

$n_2 - n_1$  dimensional system (19). We estimate the common autoregressive component using the parsimonious ARMA model on aggregates. Indeed has been previously explained that although in a VAR( $p$ ) the  $n$  implied processes are univariate ARMA( $p^*, q^*$ ), we can find a linear combination with  $\tilde{q} < q^*$ . Also Cubadda, Hecq and Palm (2007) have shown that this estimator has the best properties among the methods considered. It emerges indeed that the bias increases if we include series with an AR coefficient far from the final equation parameters but less than we might expect. In particular the difference is not so big when  $n$  increases. This goes into the direction to empirically find the series coming from the same set of final equations. Working on the serial correlation properties after estimating an autoregressive model is a first step. We can also think of applying a strategy based on using the properties of the covariance matrix of a vector MA-process to evaluate the likelihood function as we propose in Appendix 1.

## 4 Analysis of per capita real GDP in Latin America

We study the first differences of the log levels of the per capita real gross domestic product (namely the growth rates) of nine Latin American economies. We have used annual data from 1950 to 2002, i.e. 53 observations. Table 4 reports the name of these countries as well as the univariate ARMA models we have identified. As far as short-run co-movements are concerned, we can first form a group that includes Brazil, Columbia, Peru and Guatemala. Indeed, the univariate processes for the other countries are white noise.<sup>9</sup>

<sup>9</sup>Note that when looking for a set of countries sharing common cycles, white noise processes do not affect the order of the univariate ARMA processes of the remaining variables. This is because having one more white noise series included leads to consider the degree  $(n+1-(s+1))p = (n-s)p$  for the implied AR and MA parts for instance. But this simplifies the searching procedure.

Table 4: Identified ARMA(p,q) models

Identified components:	AR	MA
Brazil	1	0
Chile	0	0
Columbia	1	0
Mexico	0	0
Peru	1	0
Venezuela	0	0
Argentina	0	0
Ecuador	0	0
Guatemala	1	0

With these four countries,  $p = 2$  for the log levels was enough to capture the dynamics using a VAR. Moreover we do not conclude to the existence of long-run relationships between the log of per capita real GDP using Johansen's trace tests. Consequently, the VAR(1) for first differences implies ARMA(4,3) models for growth rates. This is not compatible with the parsimonious AR(1) we identify from the data. We interpret this as a sign of the presence of short-run co-movements. Indeed, ARMA(1,1) and AR(1) are compatible with  $s = 3$  short-run relationships. If we estimate ARMA processes for the averages of the four countries the most parsimonious model is the AR(1)

$$\Delta \bar{z}_t = \underset{(2.45)}{0.057} + \underset{(4.37)}{0.531} \Delta \bar{z}_{t-1} + \hat{e}_t$$

in which we do not see any sign of autocorrelation in  $\hat{e}_t$ .<sup>10</sup>

From these results we conjecture that there exists strong short-run co-movements characterized by a unique common cycle within that group of four countries. Both the individual ARMA orders and the absence of any autocorrelation in the residuals of the aggregated estimated equation lead to this conviction. To have an alternative look at this issue (it is possible in this case because  $n$  is small) we report in Table 5 SCCF test statistics for the number of cofeatures using a canonical correlation approach as well as the value of information criteria.  $p - val$  and  $p - val^{cor}$  refer to the  $p - value$  associated with respectively the asymptotic and the small sample corrected test statistics (Hecq, Palm and Urbain, 2006 and Hecq 2006). AIC, HQ and SC are the obvious acronyms for model selection criteria. It is shown with simulations in Hecq (2006) that both HQ and SC give a correct rank in large samples while AIC underestimates  $s$ . It emerges from Table 5 that using the four countries we cannot reject the hypothesis that there exist  $s = 3$  SCCF

<sup>10</sup>Note that the MG would produce a smaller coefficient as the Monte Carlo experiment illustrates with  $\hat{\phi}_1^{MG} = 0.431$ .

Table 5: SCCF test statistics					
	$s = 0$	$s \geq 1$	$s \geq 2$	$s \geq 3$	$s = 4$
$p\text{-value}$	-	0.54	0.23	0.06	<0.001
$p\text{-value}^{cor}$	-	0.56	0.27	0.09	<0.001
AIC	-28.974	-29.006	-29.022	-29.003	-28.729
HQ	-28.742	-28.789	-28.848	-28.902	-28.729
SC	-28.368	-28.438	-28.567	-28.738	-28.729

relationships. The normalized cofeature relationships are

$$\begin{cases} \Delta \ln Col_t - 0.523 \Delta \ln Bra_t \\ \Delta \ln Guat_t - 0.880 \Delta \ln Bra_t \\ \Delta \ln Peru_t - 0.933 \Delta \ln Bra_t \end{cases} .$$

However, our specific to general approach allows us to directly consider the series that might have a cyclical component in common. Starting with the VAR for so many variables, it would not have been possible to determine such a group and other countries could have been added (see Hecq et al. 2006 and Hecq 2006 with the same kind of data). Moreover our approach is robust to misspecification of the number of cointegrating vectors, which is not the case with usual tests. For instance in this case, the same canonical correlation tests statistics with a misspecified cointegrating rank  $r = 1$  gives  $s = 2$ .

## 5 Conclusion

This paper has proposed a new strategy to study the presence of co-movements in large set of series without the need to develop and jointly estimate a large complex multivariate model for these series. The strategy exploits the information about common dynamics present in individual ARMA estimations. Many further developments are currently under investigations such as inter alia the development of panel unit root tests and the extension to seasonal models. Moreover, the small sample and asymptotic properties must be more deeply evaluated.

The tools we introduce can be extended in further research. For instance they allow to work on co-movements between variables or on convergence among economies, to forecast series, to build indices such as business cycle components, to unravel trends from cycles,... as if we had a full parametric system with many variables. The advantages of our approach are: 1) its feasibility when it is not possible to jointly analyze a complete system or when we prefer to work using a sub-system, 2) the accuracy of forecasts, 3) the ease of its implementation when a large number of jointly dependent variables has to be studied in complex situations, 4) the potential empirical applications in many fields.

## 6 References

- AHN, S. K. AND G. C. REINSEL (1988), Nested Reduced Rank Autoregressive Models for Multiple Time Series, *Journal of the American Statistical Association* 83, 849-856.
- CUBADDA, G. AND A. HECQ (2001), On Non-Contemporaneous Short-Run Comovements, *Economics Letters*, 73, 389-397.
- CUBADDA, G. (2007), A Unifying Framework for Analyzing Common Cyclical Features in Cointegrated Time Series, Università di Roma "Tor Vergata working paper.
- CUBADDA, G., HECQ, A. AND F.C. PALM (2007), Macro-panels and Reality, Maastricht University Research Memorandum.
- ENGLE, R. F. AND S. KOZICKI (1993), Testing for Common Features (with comments), *Journal of Business and Economic Statistics* 11, 369-395.
- FORNI M., M. HALLIN, LIPPI M. AND L. REICHLIN (2000), The Generalized Dynamic Factor Model: Identification and Estimation, *Review of Economics and Statistics* 82, 4, 540-554.
- FRANSES, PH.H. (1998), *Time Series Models for Business and Economic Forecasting*, Cambridge University Press.
- GOURIÉROUX, CH. AND I. PEAUCELLE (1989), Detecting a Long-run Relationship, CEPREMAP Discussion Paper 8902.
- HECQ, A. (2006), Cointegration and Common Cyclical Features in VAR Models: Comparing Small Sample Performances of the 2-Step and Iterative Approaches, University of Maastricht Research Memorandum.
- HECQ, A., PALM, F.C. AND J.P. URBAIN (2006), Common Cyclical Features Analysis in VAR Models with Cointegration, *Journal of Econometrics* 132, 117-141.
- JOHANSEN, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models* (Oxford University Press: Oxford).
- LÜTKEPOHL, H. (2005), *New Introduction to Multiple Time Series Analysis*, (Springer Verlag: Berlin).
- LÜTKEPOHL, H. (1996), *Handbook of Matrices*, John Wiley and Sons.
- MARAVALL, A. AND A. MATHIS (1994), Encompassing Univariate Models in Multivariate Time Series: a Case Study, *Journal of Econometrics* 61, 197-233.
- PALM, F.C. (1977), On Univariate Time Series Methods and Simultaneous Equation Econometric Models, *Journal of Econometrics* 5, 379-388.
- PALM, F.C. AND A. ZELLNER (1980), Large-Sample Estimation and Testing Procedures for Dynamic Equation Systems, *Journal of Econometrics* 12, 251-283.
- QUENOUILLE, M.H. (1957, 1968 2ND ED), *The Analysis of Multiple Time-Series* (New-York, Hafner Publishing Co.).
- REINSEL, G. (1983), Some Results on Multivariate Autoregressive Index Models, *Biometrika* 70, 145-156.
- SCHLEICHER, CH. (2007), Codependence in Cointegrated Autoregressive Models, *Journal of Applied Econometrics* 22, 137-159.



- STOCK J.H. AND M.W. WATSON (2002), Macroeconomic Forecasting Using Diffusion Indexes, *Journal of Business and Economic Statistics* 20, 2, 147-162.
- TIAO, G. C. AND R. S. TSAY (1989), Model Specification in Multivariate Time Series (with comments), *Journal of Royal Statistical Society, Series B* 51, 157-213.
- VAHID, F. AND R. F. ENGLE (1993), Common Trends and Common Cycles, *Journal of Applied Econometrics* 8, 341-360.
- WALLIS, K.F. (1977), Multiple Time Series Analysis and the Final Form of Econometric Models, *Econometrica* 45, 1481-97.
- ZELLNER, A. AND F.C. PALM (1974), Time Series Analysis and Simultaneous Equation Econometric Models, *Journal of Econometrics* 2, 17-54.
- ZELLNER, A. AND F.C. PALM (1975), Time Series and Structural analysis of Monetary Models of the US Economy, *Sanhya: The Indian Journal of Statistics, Series C* 37, 12-56.
- ZELLNER, A. AND F.C. PALM (2004), *The Structural Econometric Time Series Analysis Approach*, Cambridge University Press.
- ZELLNER, A. (1962), An Efficient Method for Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias, *Journal of the American Statistical Association* 57, 348-368.

## 7 Appendix 1: The covariance matrix

In order to obtain the covariance matrix, let us construct for every series the deviation from the population autoregressive part  $w_{it} = z_{it} - \alpha_i - \sum_{j=1}^{p_i} \phi_{ij} z_{it-j}$  where we know that the  $w_{it}$  are finite order MA( $q$ ). We can denote  $\det[\Phi(L)]z_t - \alpha_i = w_t$  with then  $w_t = \Phi(L)^{adj} \varepsilon_t = H(L) \varepsilon_t$  where for every variable we have

$$w_{it} = H_{i\cdot}(L) \varepsilon_t = \theta_i(L) u_{it} \quad (20)$$

$$= u_{it} + \theta_{i1} u_{it-1} + \dots + \theta_{iq_i} u_{it-q_i}, \quad (21)$$

where  $H_{i\cdot}(L) = \Phi_{i\cdot}(L)^{adj}$  is the  $i$ th row of  $\Phi(L)^{adj}$ . We will focus on the VMA(1) model in this paper for simplicity reasons and because this is the model that we find in most empirical applications. In this case  $\Phi(L)^{adj}$  is of degree one and denoted  $\Phi_1^{adj} = H_1$  with  $w_t = \varepsilon_t + H_1 \varepsilon_{t-1}$ . Consequently for  $t \neq t'$ , variances and autocovariances are  $E(w_t w_t') = \Sigma + H_1 \Sigma H_1' \equiv D$ ,  $E(w_t w_{t-1}') = H_1 \Sigma$ ,  $E(w_{t-1} w_t') = \Sigma H_1' \equiv P$  and  $E(w_t w_{t-l}') = 0$  for  $l > 1$ . Collecting all these matrices for  $t = 1 \dots T$  gives a possibly large  $nT \times nT$  matrix

$$\mathbf{W} = \begin{pmatrix} D & P' & 0 & \dots & 0 \\ P & D & P' & \ddots & \vdots \\ 0 & P & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & P' \\ 0 & \dots & 0 & P & D \end{pmatrix}. \quad (22)$$

We do not observe nor estimate  $H_1$  in the complete VMA. Also we do not have the  $\varepsilon_t$  of the VAR but only estimates of the individual moving average parameters and error terms  $u_{it}$ . However given the equivalence (20) we propose to obtain an estimate of  $E(w w')$  using the empirical counterpart  $\hat{D} = \frac{1}{T} \Sigma \hat{w}_t \hat{w}_t'$  and  $\hat{P} = \frac{1}{T} \Sigma \hat{w}_t \hat{w}_{t-1}'$  to consistently estimate the unknown block-bands in (22) with  $\hat{w}_t$  the  $i$ -th element being  $\hat{w}_{it} = z_{it} - \hat{\alpha}_i - \sum_{j=1}^{p_i} \hat{\phi}_{ij} z_{it-j}$ .

Remains the problem of manipulating the block-band symmetric matrix  $\mathbf{W}$ , for instance to invert it or to obtain its determinant. In the special case where the autocovariances are zero (i.e.  $H_1 \Sigma = 0$ ) we have a block diagonal matrix that can be manipulated like in Zellner (1962), namely  $\det[\mathbf{W}] = \prod_{t=1}^T \det[D]$ . But more interesting is the more general VMA(1) case in which  $\mathbf{W}$  is a block-triangular symmetric matrix (22). Because  $\mathbf{W}$  is positive definite, its complete LU-decomposition  $\mathbf{W} = (\mathbf{P} + \mathbf{T}) \mathbf{T}^{-1} (\mathbf{T} + \mathbf{P}')$  exists.  $\mathbf{P} =:$  blocktridiag $\{P, 0, 0\}$ ,  $\mathbf{T} =:$  blockdiag $\{T_k\}$  and

$$T_k = \begin{cases} D, & k = 1 \\ D - P T_{k-1}^{-1} P', & k = 2 \dots T. \end{cases}.$$

This is a recursive algorithm that only involves inversion of  $n \times n$  matrices and consequently is feasible even for large  $n$  and  $T$  after  $D$  and  $P$  have been replaced by consistent estimates  $\hat{D}$  and  $\hat{P}$ . We use this decomposition to compute determinants. Indeed because  $(\mathbf{P} + \mathbf{T})$  and  $(\mathbf{T} + \mathbf{P}')$  are block-triangular and  $\mathbf{T}$

block-diagonal we have

$$\det[\mathbf{W}] = \det[\mathbf{T}] = \prod_{k=1}^T \det[T_k]. \quad (23)$$

However and although the parameters of individual ARMA models as well as the average coefficient on the aggregates are consistently estimated by ML, we cannot use traditional techniques such as the likelihood ratio or information criteria to compare  $\det[\mathbf{W}]^{res}$  and  $\det[\mathbf{W}]^{unr}$  where "res" and "unr" refer to the determinants of the covariance matrix (23) obtained respectively under restrictions on the autoregressive component or free from these restrictions. A theoretical reason is that a likelihood ratio for the multivariate model with parameters evaluated at consistent but not necessary efficient system estimates it is not a maximum likelihood ratio. Therefore, its asymptotic distribution has to be determined under the null. Solutions are the use of system maximum likelihood or of efficient two-step estimators of the VARMA as proposed in Palm and Zellner (1980). However, not only is this not feasible for large  $n$ , implementing that approach is rather intricate.

Let us consider a special case and denote by  $\det[\mathbf{W}]^{ind}$  the determinant of the covariance matrix obtained when the residuals are constructed from individual estimation and  $\det[\mathbf{W}]^{av}$  when the residuals  $\hat{w}_{it} = z_{it} - \hat{\alpha}_i - \sum_{j=1}^{p_i} \hat{\phi}_j^{av} z_{it-j}$  are taken using common coefficient estimates from the averaged series. The two models will most of the time be non-nested because it can happens that a fraction of the  $n$  series are ARMA(1,1) and the model for the aggregates is an AR(1). Consequently  $\det[\mathbf{W}]^{ind}$  is not necessarily smaller than  $\det[\mathbf{W}]^{av}$ . Also the use of information criteria is useless because the difference between these two determinants might be very large so that the adjustment by the number of parameters has no effect.

To illustrate this point let us compute the frequency with which  $\det[\mathbf{W}]^{ind}$  is smaller than  $\det[\mathbf{W}]^{av}$  by means of a small Monte Carlo experiment and compute the frequency with which  $\Upsilon = (\det[\mathbf{W}]^{av} - \det[\mathbf{W}]^{ind}) \leq 0$ . If individual estimations are close to each other and close to the estimate on aggregates, then the proportion  $\Upsilon$  should be different from 50%. To illustrate this hypothesis we simply use a sign test on 1000 replications from (18) in which we both estimate each equation separately and on the aggregates using  $S = \sum_{i=1}^{1000} u(\Upsilon)$ , where  $u(\Upsilon) = 1$  if  $\Upsilon \leq 0$  and  $u(\Upsilon) = 0$  if  $\Upsilon > 0$ . The standardized statistics is

$$\bar{S} = \frac{(S - 0.5 \times 1000)}{0.5\sqrt{1000}} \sim N(0, 1)$$

Table 6 reports the frequency  $S/1000$  as well as the value of the statistics  $\bar{S}$  in two different settings for  $\det[\mathbf{W}]^{ind}$ . In the first case each series is taken from the true parameter  $\phi_{i1} = 0.5$  for  $i = 1 \dots n$ . In the second case coefficients of  $n$  ARMA(1,1) models are estimated and we have  $\phi_{i1} = \hat{\phi}_{i1}$  for  $i = 1 \dots n$ . In both of these two situations, the coefficient on the averaged series is obtained using the best parsimonious ARMA on the aggregates. It emerges from Table 6 that the frequency with which one model is preferred is roughly 0.5 in most cases. But there are significant deviations for small  $n$  ( $n = 5$ ) when the coefficients are estimated from the individual series. Obviously, in practice we do not have many times the same process in order to compute  $\bar{S}$ . We leave for further investigation a bootstrap test on this issue.

Table 6: Frequencies with which aggregates coefficient model is chosen and sign test

$n/T$		$\frac{S}{1000}$			$\bar{S}$		
		50	100	250	50	100	250
5	$\phi_{i1} = 0.5$ vs $\hat{\phi}_1$	0.505	0.514	0.509	0.31	0.88	0.56
11		0.507	0.478	0.510	0.44	-1.39	0.63
23		0.521	0.532	0.483	1.32	2.02	-1.07
5	$\hat{\phi}_{i1}$ vs $\hat{\phi}_1$	0.588	0.594	0.590	5.56	5.94	5.69
11		0.489	0.483	0.561	-0.69	-1.07	3.85
23		0.477	0.487	0.506	-1.45	-0.82	-0.37